Diabetic Complications Consortium

Application Title: Iterative Integration of Single Cell RNA-seq and ATAC-seq Data for Inferring Transcription Factor Regulatory Networks in Macrophages

Principal Investigator: Yang Dai/Timothy Koh (MPI)

1. Project Accomplishments

Macrophages are capable of performing a diverse array of tasks during wound healing, ranging from destructive killing functions to pro-healing and homeostatic duties. Recent studies, especially those based on the single-cell RNA-seq approach, have demonstrated that macrophages adopt a spectrum of phenotypes during skin wound healing in healthy and diabetic mice[1-3]. To identify the transcriptional regulatory mechanisms potentially driving the heterogeneity in macrophage phenotypes during the healing of skin diabetic wounds, in the first funding year, we applied our recently developed tool to the time-course scRNAseq datasets obtained from the skin wounds of the on-diabetic and diabetic db/db mice on days 3, 6, and 10 post-injury to infer the activity of transcription factors (TF) in individual cells using the BITFAM tool[4]. BITFAM leverages all published TF ChIP-seq data in scRNAseq data analysis to predict the activities of TFs in individual cells and provides a rank list of targeted genes for each TF. We have predicted the TFs with differential activity in each macrophage subpopulation in non-diabetic and diabetic mice and estimated the activities of the BITFAM inferred gene targets using AUCell scores[5]. During the second year, we applied BITFAM using the predicted TF-targets from the scATACseq data for context-specific TF binding sites and the target genes. The scATACseq data were generated from Dr. Koh's lab under R35GM136228. Our analysis results reviewed distinct transcription factor activity profiles in macrophage subpopulations, suggesting novel testable hypotheses on understanding the key regulators in skin wound healing.

2. Specific Aim

Specific Aim 1B. Infer TF activity using the scRNAseq data of macrophages in skin wounds obtained from non-diabetic (GSE154400)[1] day 3, 6, and 10; and scATACseq data of non-diabetic mice on days 3 and 6 (data unpublished, the scATACseq data at Day 10 was unavailable due to experimental failure.)

Table 1 Number of cells obtained

	scRANseq	**scATACseq**
D3	447	7052
D6	292	6990
Total	739	14042

Results:

First, we identified distinct macrophage subpopulations from the scRNAseq and scATAseq data data (Table 1) obtained from Dr. Koh's lab using Seurat (ver3)[6]. The analysis revealed 5 major

macrophage subpopulations (Fig.1). Noticeably, cells in Cluster 0 were persistently observed in both Day3 and Day3 and with marker genes indicating characteristic antigen-presenting cells.

Second, we obtained the open chromatin regions from the scATACseq data and identified overrepresenting motifs of TFs, which are specific to macrophage skin wounds. Then we retrieved targets of these TFs from the ChIP-Atlas database[7] using ChIPseq experiments on macrophages and skins. The targeted genes of a TF were determined based on the TF binding sites located at upstream and downstream of transcription start sites.

Third, we combined the TF-gene targets with the default database of BITFAM to generate the TF activity profiles for all cells obtained from the scRNAseq data. Then we performed the differential analysis of the activity and identified TFs specific to each macrophage subpopulation. For each of the TF, the activity of their target genes was calculated using AUCell[5] (Fig. 2). Fig.3 shows examples of analysis results of the several previously known members of the TF family cebp.

We are currently preparing a manuscript about comprehensive analysis results on transcription factor activity in macrophage subpopulations.

3. Publication

1. Zandigohar M and Dai Y. "Information retrieval in single cell chromatin analysis using TF-IDF transformation methods". *The Proc. of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 877-882. 2022. doi:10.1109/BIBM55620.2022.9994949

References

1. Pang, J., Maienschein-Cline, M., and Koh, T.J. (2021). Enhanced Proliferation of Ly6C(+) Monocytes/Macrophages Contributes to Chronic Inflammation in Skin Wounds of Diabetic Mice. J Immunol *206*, 621-630. 10.4049/jimmunol.2000935.
2. Pang, J., Maienschein-Cline, M., and Koh, T.J. (2022). Monocyte/Macrophage Heterogeneity during Skin Wound Healing in Mice. J Immunol *209*, 1999-2011. 10.4049/jimmunol.2200365.
3. Pang, J., Maienschein-Cline, M., and Koh, T.J. (2021). Reduced apoptosis of monocytes and macrophages is associated with their persistence in wounds of diabetic mice. Cytokine *142*, 155516. 10.1016/j.cyto.2021.155516.
4. Gao, S., Dai, Y., and Rehman, J. (2021). A Bayesian inference transcription factor activity model for the analysis of single-cell transcriptomes. Genome Res *31*, 1296-1311. 10.1101/gr.265595.120.
5. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nature methods *14*, 1083-1086. 10.1038/nmeth.4463.
6. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821. 10.1016/j.cell.2019.05.031.

7. Zou, Z., Ohta, T., Miura, F., and Oki, S. (2022). ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. Nucleic Acids Research *50*, W175-W182. 10.1093/nar/gkac199.

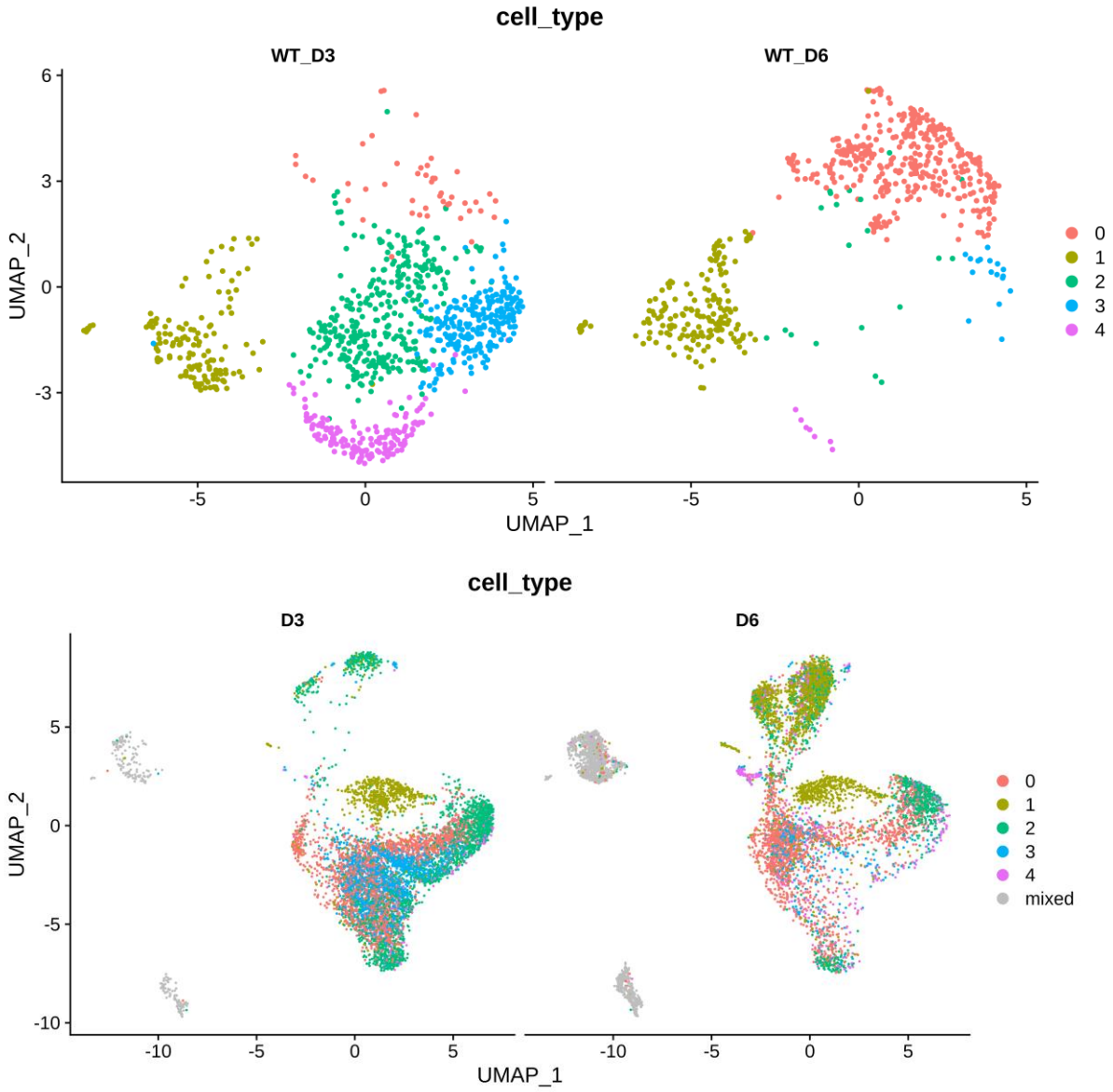

Fig.1 The clustering of the macrophage subpopulations. Left: macrophage subpopulations identified from the scRNAseq data at D3 and D6 for wild-type mice; Right: macrophage subpopulations identified from the scATACseq data at D3 and D6 for wild-type mice. Clustering labels were transferred from the scRNAseq data based on the gene score profiles using Seurat.

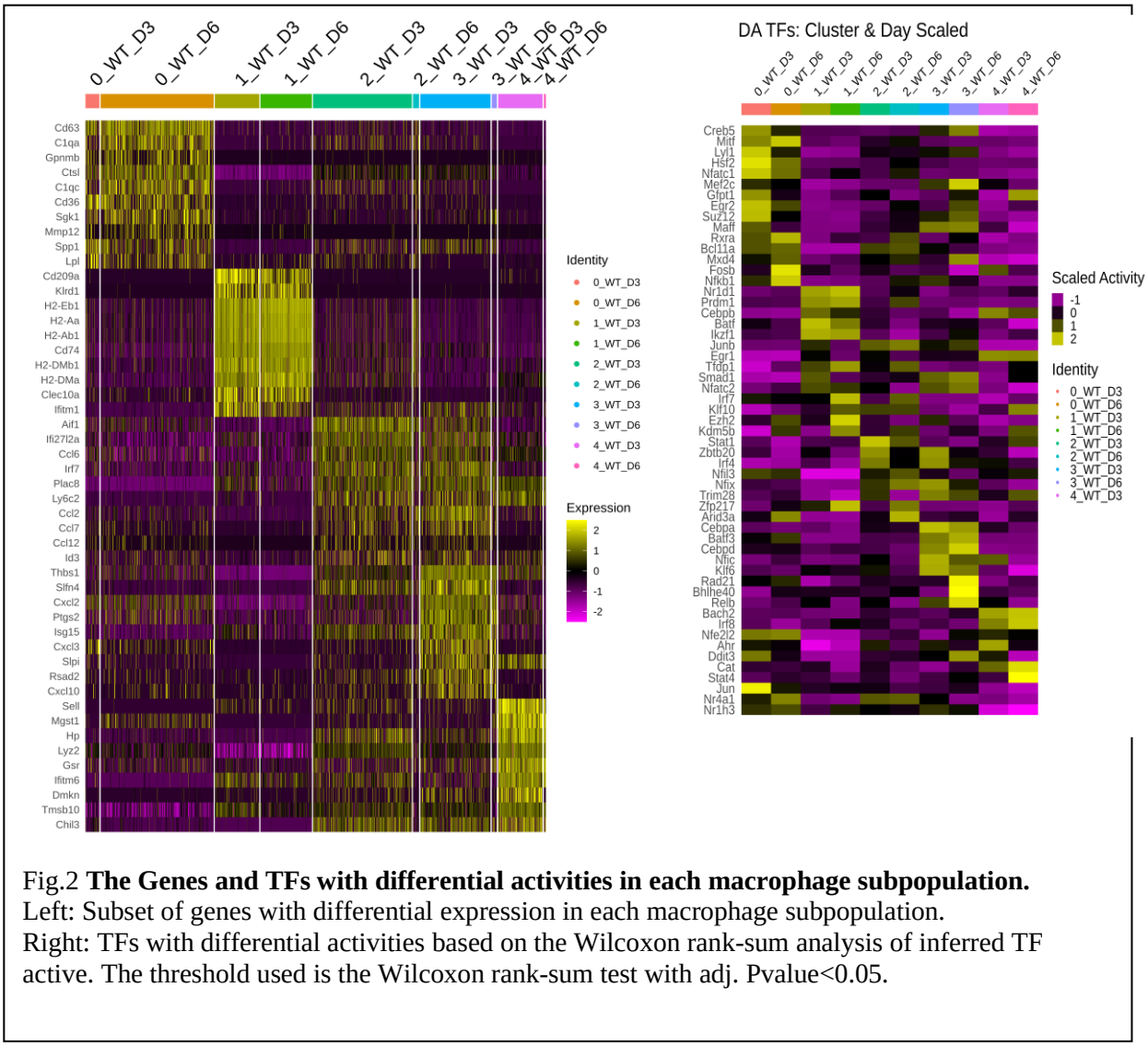

Fig.2 **The Genes and TFs with differential activities in each macrophage subpopulation.** Left: Subset of genes with differential expression in each macrophage subpopulation. Right: TFs with differential activities based on the Wilcoxon rank-sum analysis of inferred TF active. The threshold used is the Wilcoxon rank-sum test with adj. Pvalue<0.05.

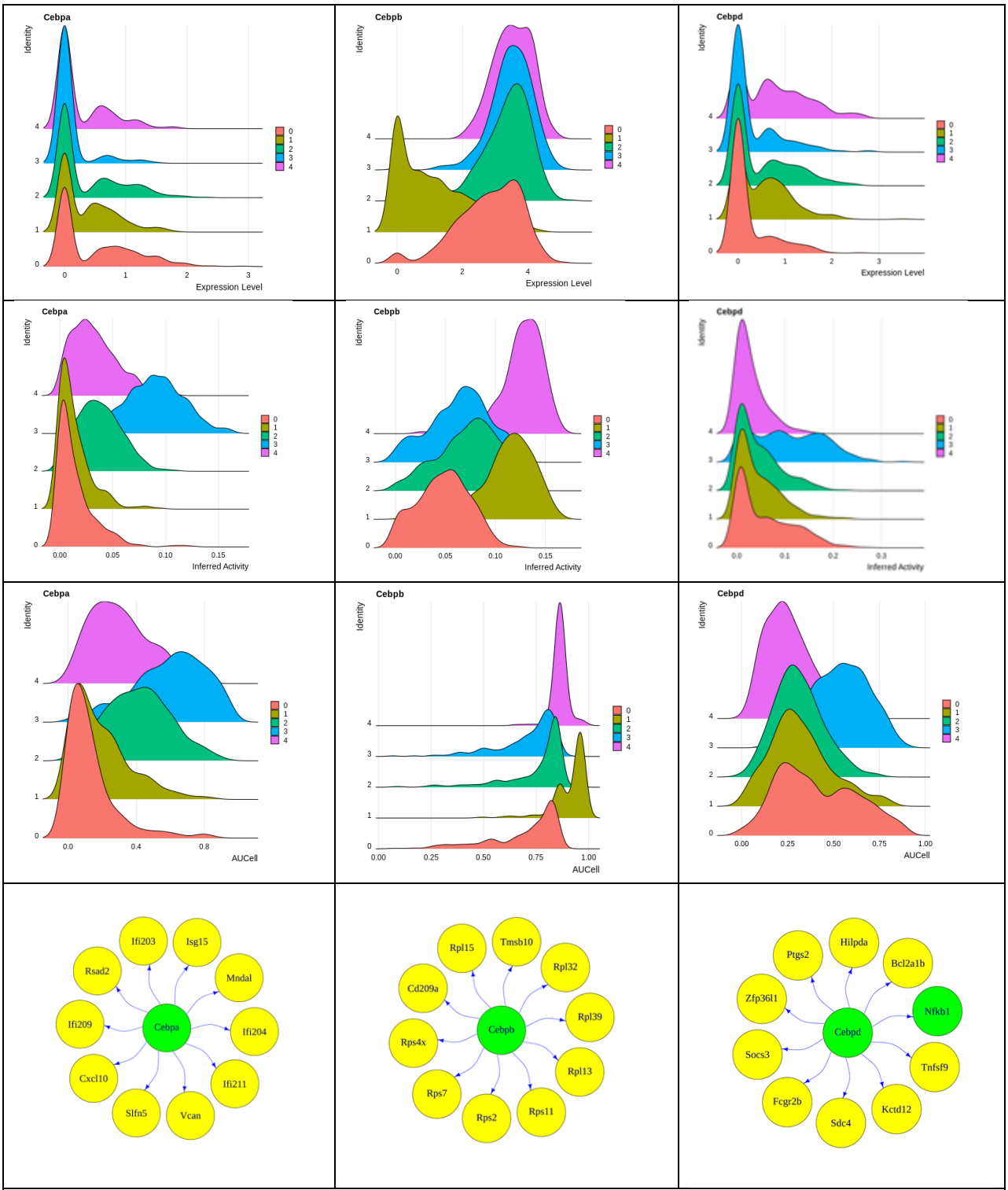

Fig. 3. Predicted activity and target genes of the transcription family cebp in macrophage subpopulations.
Row 1**:** Gene expression of cebpa, cebpb and cebpd in macrophage subpopulations;
Row 2: Inferred activity of transcription factors, Cebpa, Cebpb, and Cebpd in macrophage populations;
Row 3: The AUCell scores of the top 10 target genes of Cebpa, Cebpb, and Cebpd in the macrophage subpopulations;
Row 4: Top 10 target genes of Cebpa, Cebpb, and Cebpd.